Diversity Arrays Technology effectively reveals DNA polymorphism in a large and complex genome of sugarcane

Katarzyna Heller-Uszynska · Grzegorz Uszynski · Eric Huttner ·
Margaret Evers · Jason Carlig · Vanessa Caig · Karen Aitken ·
Phillip Jackson · George Piperidis · Mike Cox · Ross Gilmour ·
Angelique D'Hont · Mike Butterfield · Jean-Christophe Glaszmann ·
Andrzej Kilian

Abstract Diversity Arrays Technology (DArT) provides whole genome profiling for hundreds to thousands of polymorphic markers in a single assay using a high-throughput microarray platform. The presented work aimed to establish DArT genotyping for the genetically challenging genome of sugarcane. Due to the genome complexity of this sugar-producing crop of high economic importance, an application of DArT genotyping to this species required extensive testing and optimization. As the method of genome complexity reduction determines the efficiency of polymorphism identification in DArT, various approaches and several methods were tested, in order to establish the most optimal. The sugarcane DArT markers generated with these established methods identified high genetic differentiation of sugarcane ancestral species from modern cultivars, in agreement with the data available for other types of molecular markers for this crop. The majority of sugarcane DArT markers segregated in a Mendelian fashion and were readily incorporated into the framework genetic map. As the DArT markers are sequence-ready genomic clones, we sequenced 384 clones and found that one-third of sequenced markers came from the transcribed portion of the sugarcane genome. The presented results further validate the potential of DArT technology in providing cost-effective genetic profiles for plants, irrespective of their genome complexity, for effective applications in molecular-assisted breeding, diversity analysis or genetic identity testing.

Keywords DArT genotyping · Sugarcane · Molecular markers · Microarray · Whole genome profiling

K. Heller-Uszynska · G. Uszynski · E. Huttner ·
M. Evers · J. Carlig · V. Caig · A. Kilian (✉)
Diversity Arrays Technology Pty. Ltd, Canberra,
ACT 2600, Australia
e-mail: a.kilian@diversityarrays.com

K. Aitken · P. Jackson
CSIRO Plant Industry, St. Lucia, QLD 4067, Australia

G. Piperidis · M. Cox · R. Gilmour
BSES Limited, Brisbane, QLD 4068, Australia

A. D'Hont · J.-C. Glaszmann
CIRAD, UMR 1096 DAP, Montpellier, France

M. Butterfield
SASRI, Mt Edgecombe, South Africa

Introduction

Sugarcane is an important industrial crop cultivated for its sucrose which accumulates in the plant stem internodes. This clonally propagated large grass belonging to the Poaceae family is also characterized

by high content of valuable chemicals, very high photosynthetic efficiency and one of the highest efficiencies of water usage among plants. These traits provide a very attractive option of using sugarcane as a biomass producer and biofactory of desirable compounds (Botha 2006). Currently, sugarcane is grown on 13 million hectares in more than 100 countries across tropical and subtropical zones. Of almost 150 million tons of sugar consumed worldwide in 2005, 74% was made from sugarcane cultivated primarily in Brazil and India (FAO/FAO-STAT website). The world demand for sugarcane is growing rapidly in recent years due to increasing interest in utilising sugarcane sucrose to produce bioethanol as a biofuel alternative to petrol. Additionally, sugarcane's efficient biomass production attracts a lot of interest for its potential to be converted to electricity as a response to the requirement to limit the world's dependence on fossil fuels and to decrease greenhouse gas emissions (Goldemberg 2007).

Improving the efficiency of sugarcane breeding programs is an important component in addressing this growing demand for sugarcane. The conventional improvement of sugarcane varieties takes on average 12 years from an initial cross to the release of improved cultivars (Silva and Bressiani 2005). It is anticipated that sugarcane molecular genetics tools can accelerate the rate of genetic gain in breeding programs. However, sugarcane, as one of the most genetically complex crops, presents a considerable challenge for genetic and molecular analysis. The large genome (2C) of species belonging to the *Saccharum* genus (7,440 Mbp for *S. officinarum* L., Grivet and Arruda 2001; 10,000 Mbp for a modern cultivar R570, D'Hont 2005) is characterized by a high degree of polyploidy and frequent aneuploidy (D'Hont et al. 1996). Most sugar-producing modern cultivars are derived from crosses between *S. officinarum* ($x = 10$, $2n = 8x = 80$) and *S. spontaneum* L. ($x = 8$, $2n = 5–16x = 40–128$), both highly polyploid, resulting in a high number of chromosomes (100–130) which pair and recombine in many combinations (Jannoo et al. 2004). The genome of current cultivars is derived in ~70 to 80% from *S. officinarum*, in ~10 to 20% from *S. spontaneum*, while about 10% are recombinant chromosomes of these two species (D'Hont et al. 1996).

In spite of this genome complexity, a significant progress has been made in genomics of sugarcane, including advances in molecular cytogenetics, diversity analysis, and genetic mapping (reviewed in Grivet and Arruda 2001; D'Hont et al. 2008). Additionally, a sugarcane Bacterial Artificial Chromosome (BAC) library (Tomkins et al. 1999) and over 237,000 sugarcane Expressed Sequence Tags (EST) (Vettore et al. 2006) have become available. Several types of molecular markers have been established for sugarcane including RFLPs (Da Silva et al. 1993), RAPDs (Mudge et al. 1996), SSRs (Cordeiro et al. 1999) and AFLPs (Hoarau et al. 2001). Markers developed so far have been used for molecular investigation of the sugarcane genome structure (Grivet et al. 1996; Jannoo et al. 1999a) and for tracking genetic diversity and relationships between species and cultivars (Jannoo et al. 1999b; Lima et al. 2002; Aitken et al. 2006). Genetic maps have also been constructed for *S. spontaneum*, *S. officinarum* and modern cultivars. The first genetic maps were generated with one type of marker, such as RFLPs (Grivet et al. 1996), RAPDs (Mudge et al. 1996) and more recently AFLPs (Hoarau et al. 2001), providing a genome coverage ranging from 2,008 cM for RFLPs to 5,849 cM for AFLPs. To expand genome coverage and to improve identification of linkage groups, various types of markers have been integrated. Recently, a combination of AFLP, RAF and SSR markers revealed 136 linkage groups with a total map length of 9,058 cM in the Q165 Australian cultivar using a segregating F1 population (Aitken et al. 2005).

A significant factor limiting the application of established molecular markers in breeding programs is the speed and cost at which sufficient numbers of markers can be screened across large populations of genotypes. Moreover, some PCR-based marker systems such as SNPs or SSRs rely on sequence information for their initial development, limiting their use to the species with sufficient sequence information. This especially affects polyploid species, sequencing of which is technically difficult. As modern sugarcane cultivars have approximately 110 chromosomes corresponding to a genetic map of about 17,000 cM, and given the species' high ploidy level, 5,000–7,000 markers would be needed to achieve an adequate coverage for genetic mapping and breeding applications. We therefore attempted to

develop Diversity Arrays Technology (Jaccoud et al. 2001) for this crop. DArT combines the low cost and high throughput of the DNA microarray platform with the ability to identify various types of DNA polymorphisms (INDELs, SNPs, methylation changes) (Kilian et al. 2005). Hundreds to thousands (depending on the polymorphism level) of polymorphic features (DArT markers) distributed across the genome are discovered in a single hybridisation-based assay with an accuracy of 99.8% (Wenzl et al. 2004; Akbari et al. 2006). DArT markers do not require further assay development once they are discovered and they are sequence-independent but sequence-ready. The DArT technology proof-of-concept was originally developed on rice (Jaccoud et al. 2001). Subsequently, the technology was established for plant species with bigger and more complex genomes such as barley (Wenzl et al. 2004), polyploid wheat (Akbari et al. 2006) and polyploid oats (Tinker et al. 2009) as well as for another fifty plant species, two species of plant pathogenic fungi, a bacterium and two animal species (Lezar et al. 2004; Wittenberg et al. 2005; Xia et al. 2005; Yang et al. 2006; Sessittsch et al. 2006; Mace et al. 2008; James et al. 2008; Bonin et al. 2008; Mantovani et al. 2008). An increasing number of published studies indicate a successful deployment of DArT markers in a variety of applications, for example construction of high-density consensus genetic maps providing a framework for transferring genetic information between different marker systems, identification of quantitative trait loci (QTL) for variety of phenotypic traits and studies of genetic diversity in crops such as wheat, barley, sorghum, oats and others, as well as genomic evolutionary studies in non-model organisms (www.DiversityArrays.com/publications).

Here, we describe the effective development of DArT technology for sugarcane. Numerous methods of genome complexity reduction were evaluated for their efficiency in detecting DArT markers. The best methods were established and successfully tested in diversity analysis of sugarcane ancestral accessions and modern cultivars. The majority of identified sugarcane DArT markers segregated in a Mendelian fashion and were incorporated into a framework genetic map. The successful validation of DArT genotyping in sugarcane demonstrates that the technology can serve as a high-throughput genome scan for plants with high levels of genome complexity,

Table 1 Sugarcane genotypes used for testing the methods of genome complexity reduction and developing DArT arrays

Genotype name	Classification
Q124	Cultivar
F172	Old cultivar and parent
Korpi	*Saccharum officinarum* L.[a]
POJ 2878	Old cultivar and important breeding ancestor
Badila	*Saccharum officinarum* L.[a]
Mandalay	*Saccharum spontaneum* L.[a]
Mida	Cultivar
QN58-829	Elite parent
Q208	Cultivar
QN67-3184	Elite parent
CP74-2005	Cultivar and elite parent
Co475	Old cultivar and important breeding ancestor
Tabongo	*Saccharum spontaneum* L.[a]
Q201	Cultivar
Trojan	Old cultivar and important breeding ancestor
Tellus	Cultivar

[a] Indicates ancestral species

cost-effectively providing enough markers for a range of applications such as whole-genome-based molecular breeding, diversity analysis or genetic identity testing for a broad spectrum of plant species.

Materials and methods

Plant material

Leaf samples were supplied by the Cooperative Research Centre for Sugar Industry Innovation through Biotechnology (CRCSIIB, Australia). Table 1 lists 16 sugarcane genotypes used for testing of the genome complexity reduction methods, developing DArT libraries and arrays, and generating genomic representations to test performance of the constructed arrays.

The Q813 (*S. officinarum*) and a cultivar F172 were used to estimate reproducibility and robustness of a *PstI/TaqI* genome complexity reduction method.

The Q813 was a driver in the construction of the *PstI/TaqI* Suppression Subtractive Hybridization (SSH) DArT library, while a tester was a mixture of 16 sugarcane genotypes listed in Table 1.

The IJ76-514 (*S. officinarum*) and Q165 (an Australian cultivar), parents of a mapping

population, were used to develop the *PstI/TaqI* SSH DArT library constructed for an assessment of sugarcane DArT markers segregation patterns in the genotyping experiment of the 94 progeny clones and the parents of this mapping population.

DNA extraction

Genomic DNA was extracted using a modified cetyl-trimethylammonium bromide (CTAB)/chloroform/isoamylalcohol method (Doyle and Doyle 1987).

Testing the methods of genome complexity reduction

Various combinations of 12 restriction enzymes were tested in a total of 10 methods (Electronic Supplementary Material: Online Resource 1) using 16 sugarcane genotypes representing the *Saccharum* genetic diversity (Table 1). For each method, ∼100 ng of DNA of each sample listed in Table 1 was used. Digestion and ligation reactions were carried out simultaneously with 2 units of each of restriction enzymes (NEB), 80 units of T4 DNA ligase (NEB) and 0.05 μM of adaptor(s) (Online Resource 1, or specified individually). One microlitre of the digestion/ligation product was used as template for amplification with 2 units of *RedTaq* DNA polymerase (Sigma) and 0.4 μM of primer if one primer was used or 0.2 μM of primer if two primers were used. The eight methods are summarized in Online Resource 1a.

For method 9, 50 μl of *PstI/TaqI* amplicons generated as described for method 1 (Online Resource 1a) were redigested with 1 unit each of *DpnII, HpyCH4*IV*, Mse*I and *Nla*III (NEB) restriction enzymes to further decrease the genome complexity. A 25 μl digestion product was ligated to 0.05 μM of equimolar mixture of two sets of adaptors (sets A and B, Online Resource 1b) with 160 units of T4 DNA ligase (NEB). Five microlitres of the resulting ligation product was used as a template for amplification with 2 units of *RedTaq* DNA polymerase (Sigma); primers and amplification conditions used were as specified in Online Resource 1b.

Finally, a method based on the display of Miniature Inverted-repeat Transposable Elements (MITEs) was tested (method 10, Online Resource 1c). In this

method the genomic representation comprised the fragments amplified between the Terminal Inverted Repeats (TIRs) present at both ends of MITEs and an adaptor ligated to overhangs created by *Bsp1286*I restriction enzyme (S. Patarapuwadol, personal communication). Restriction with *Bsp1286*I (NEB) with simultaneous ligation of *Bsp1286*I adaptors (Online Resource 1c) was carried out as described for other methods. A 0.5 μl aliquot of digestion/ligation product was used as template for the first round of amplification with 0.04 μM of primer I (Online Resource 1c) complementary to adaptor ligated to the *Bsp1286I* site, 0.4 μM of each of the primers (II and III, Online Resource 1c) amplifying from TIRs and 1 unit of *RedTaq* DNA polymerase (Sigma). Amplification conditions were as follows: 94°C for 1 min, 15 cycles of 94°C for 30 s, 50°C for 40 s, followed by 72°C for 1 min. A 0.5 μl aliquot of the resulting amplification product served as a template for the second round of amplification performed as in the first round except that the concentration of primer I was increased tenfold and 10 more cycles were added.

An approach to enrich for polymorphic clones

In an attempt to enrich for polymorphic clones, a Suppression Subtractive Hybridization (SSH) technique was tested. The guidelines for SSH were as published by Diatchenko et al. (1996). Sugarcane ancestral species Q813 (*S. officinarum*) was used as a driver to subtract the shared component of the genome from a mix of 16 genotypes used as a tester (listed in Table 1). At first, the genome complexity of driver and tester was reduced as described above for method 9 (Online Resource 1b). Subsequently, driver and tester DNA was extracted with phenol/chloroform and precipitated with isopropanol according to the standard methods (Sambrook et al. 1989). Two hundred nanograms of purified tester DNA was ligated to 10 μM of two sets of adaptors (set A and set B, Online Resource 1b) in two separate ligation reactions, each reaction with 80 units of T4 DNA ligase (NEB). Subtraction was done in two quantitative ratios of driver to tester (30:1 and 10:1) and carried out in one and two rounds of subtractive hybridization. For one round subtraction, DNA quantities of driver and tester appropriate to the specified ratios were mixed separately for each of the

two sets of adaptors. The mixed DNA was precipitated with isopropanol, dissolved in 1.5 μl Subtraction Hybridization Buffer (Diatchenko et al. 1996), overlaid with 5 μl mineral oil, denatured at 98°C for 1.5 min and allowed to hybridize for 5 h at 68°C. Once the hybridization was completed, the subtraction product was diluted with 200 μl Subtraction Dilution Buffer (Diatchenko et al. 1996), and then heated at 72°C for 7 min. The two rounds subtraction was done as one round but after 5 h incubation at 68°C, the hybridization products with set A of adaptors were mixed with hybridization products with set B of adaptors, separately for different ratios. Subsequently, 300 ng of freshly denatured (98°C for 1.5 min) driver was added to each ratio variant. The mixtures were allowed to hybridize for an additional 5 h at 68°C. Once the hybridization was completed, each subtraction product was diluted with 200 μl Subtraction Dilution Buffer, and then heated at 72°C for 7 min.

One, 5 or 10 μl aliquots of each variant of the hybridization product were amplified with 0.2 μM of each of primers 5′-CTGAGTAGTGCCAGAACGG TC-3′ and 5′-TCGTAGACTGCGTATCCG-3′ and 2 units of *RedTaq* DNA polymerase (Sigma). The reaction conditions were as follows: 72°C for 7 min, 95°C for 3 min, 30 cycles of 94°C for 20 s, 58°C for 40 s, 72°C for 1 min followed by 72°C for 7 min.

The steps of DArT technology downstream of the genome complexity reduction method were as previously published (Jaccoud et al. 2001; Wenzl et al. 2004; Akbari et al. 2006) with some modifications and will be described briefly.

Preparations of DArT arrays

Each method of the genome complexity reduction was used to build a library comprising from 768 to 4,608 individual clones. For each method, genomic representations of the 16 genotypes listed in Table 1 were mixed and cloned using a "TOPO TA Cloning" kit (Invitrogen) according to the methods recommended by the manufacturer. Individual recombinant colonies were grown overnight in 384-well plates in a "freezing medium" (LB with 100 μg/ml ampicillin and mix of salts eliminating LB inhibitory effects on subsequent PCR amplification; unpublished observation). A 0.5 μl aliquot of bacterial culture was used as a template for insert amplification with 0.2 μM of

each M13 Forward and M13 Reverse primers (Invitrogen). The cycling conditions were as follows: 95°C for 4 min, 57°C for 35 s, 72°C for 1 min followed by 35 cycles of 94°C for 35 s, 52°C for 35 s and 72°C for 1 min and finally 72°C for 7 min. The amplicons were dried, washed with 70% ethanol and dissolved in a spotting buffer developed especially for poly-L-lysine-coated microarray slides (P. Wenzl, personal communication). The amplified clones suspended in the spotting buffer were printed in duplicate on the poly-L-lysine-coated slides (Erie Scientific) using a MicroGridII arrayer (Biorobotics). After printing, the DNA deposited onto slides was denatured by incubation in hot water (95°C) for 2 min, followed by dipping in MQ water supplemented with 0.1 mM DTT and 0.1 mM EDTA and drying by centrifugation at 500g for 7 min.

Genotyping using DArT arrays

Preparation of genomic representations of individual sugarcane genotypes (=targets) for hybridization to the DArT arrays was done using the same methods as for target preparation for DArT arrays generation. Genetic representations were precipitated with isopropanol and dissolved in molecular grade water (Sigma) before labelling. Denatured targets were labelled with 2.5 units of exo-Klenow fragment of *E. coli* Polymerase I (NEB), 25 μM random decamers and 2.5 nmoles of either Cy3-dUTP or Cy5-dUTP (Amersham Bioscience). Labelling was carried out for at least 3 h at 37°C. Labelled targets were denatured for 2 min at 95°C, mixed with 60 μl of hybridization solution and deposited onto the microarrays. The hybridization solution was composed of a 50:5:1 mixture of, respectively, ExpressHyb (Clonetech), herring sperm DNA (Promega), FAM-labelled polylinker of pCR 2.1 vector used for cloning (Invitrogen), and 2 mM EDTA (pH 8.0). Hybridization was carried out for 18 h at 60°C. Following hybridization, the slides were washed at room temperature in four solutions with increasing salt stringency (solution 1—1 × SSC, 0.1% SDS; solution 2—1 × SSC; solution 3—0.2 × SSC; solution 4—0.02 × SSC) and dried by centrifugation at 500g for 7 min. Finally, the microarrays were scanned using a Tecan LS300 (Grödig, Austria) confocal laser scanner. Image analysis and polymorphism identification and scoring was done with the programme DArTSoft

(version 7.4.3) specifically developed for that purpose by Diversity Arrays Technology Pty. Ltd. (www.DiversityArrays.com/software.html) and as described in Wenzl et al. (2004) and Akbari et al. (2006). Briefly, DArTSoft analyzed the microarray images, extracted the hybridization intensity data from the hybridized arrays, identified and scored polymorphism and calculated a range of quality parameters for each marker. The polymorphism analysis was performed with DArTSoft default settings which select markers of high quality and reproducibility based on extensive experience with several other genomes (Wenzl et al. 2004; Wittenberg et al. 2005; Xia et al. 2005), including polyploid wheat (Akbari et al. 2006).

Evaluation of the genome complexity reduction methods

The genetic representations of 16 samples representative of the *Saccharum* diversity and used to build the arrays (Table 1) were hybridized to the microarrays printed for each of the tested complexity reduction methods. Each sample was analyzed with replication to allow evaluation of technical reproducibility. DArTSoft analysis with the default settings was run for each hybridization experiment/method of genome complexity reduction. The most important criterion taken into consideration for choosing the best genome complexity reduction methods was the frequency of polymorphic markers. The average Call Rate (the percentage of successful allelic assignments), the average value of function "P" (measuring the level of relative target signal bimodality/polymorphism for a particular spot) and reproducibility of scores obtained from independent technical replicates were considered as supporting selection criteria.

Validation of the *Pst*I/*Taq*I method in reducing the genome complexity and testing method reproducibility

To validate the *Pst*I/*Taq*I method in reducing the genome complexity, a hybridization experiment of genomic DNA of Tabongo (*S. spontaneum*), Badila (*S. officinarum*) and Q124 (cultivar) was carried out on a *Pst*I/*Taq*I array containing 1,536 clones. Each sample was hybridized in two replicates. One hundred nanograms of genomic DNA was labeled

with Cy3-dUTP, hybridized and analyzed as described in the section "Genotyping using the DArT arrays".

In order to assess reproducibility of the *Pst*I/*Taq*I method, a genotyping experiment on two sugarcane accessions (Q813 and F172) was carried out on the *Pst*I/*Taq*I array with 1,536 clones printed in triplicate. For each genotype, two independent digestion/ligation steps were performed and from each reaction six targets were prepared using two different methods of amplification in triplicate. One amplification method used a single 50 µl volume reaction while the second method was based on bulking three independent reactions with 17 µl each. The targets were labeled with Cy3-dUTP, hybridized and analyzed as described in the section "Genotyping using the DArT arrays".

The reproducibility of replicated targets was estimated by calculating a correlation (R^2) between the genomic representations' relative signal intensities ($\log_2$ of ratio derived from fluorescence signal of genomic representation divided by reference signal).

Assessment of Mendelian behavior of DArT markers

Construction of a DArT array for the assessment of Mendelian behavior of DArT markers

The libraries built during testing of the methods of genome complexity reduction were constructed using germplasm comprising sugarcane genetic diversity, not the best suited for the characterization of DArT markers' segregation patterns in a specific genetic mapping population. In order to enrich the clone pool for clones segregating in the specific progeny, the library was constructed from the parents of the IJ76-514 (*S. officinarum*) × Q165 (an Australian cultivar) mapping population. We used the Suppression Subtractive Hybridization technique as described in the section "An approach to enrich for polymorphic clones". Each parent was used both as a tester and a driver, respectively, in a 30:1 ratio of driver to tester in one or two rounds of subtraction. The constructed library comprised 3,072 clones, 1,536 clones each for one-round and two-rounds subtraction methods. The final array used in the experiment to assess the Mendelian behavior of DArT markers consisted of 6,144 clones. The array composition is shown in Online Resource 2.

Marker segregation

Each marker was scored as present (1) or absent (0) in the 94 progenies of the IJ76-514 × Q165 mapping population. A marker was considered reliable if scored successfully in at least 75% of individuals, a threshold significantly lower compared to the thresholds used for species with less complex genomes. Markers were tested for deviation from the segregation ratios expected for single-dose (1:1) and biparental single-dose markers or double-dose markers present in only one parent (3:1) by the Chi-square (χ^2) test and declared as significantly distorted at $P \leq 0.05$.

Sequence analysis of the sugarcane candidate polymorphic DArT markers

The 384 polymorphic clones were chosen from the hybridization experiments performed using the *Pst*I/*Taq*I/SSH and *Pst*I/*Taq*I arrays. The criteria for choosing the candidate polymorphic markers were as follows: (1) if the given sample was hybridized in more than one experiment, the obtained marker scores were compared and markers with up to one discordant score were chosen; (2) if the sample was hybridized in one experiment only and thus marker score comparison could not be carried out, markers with 100% reproducibility and $P > 80$ were chosen.

Altogether, 332 *Pst*I/*Taq*I/SSH clones and 52 *Pst*I/*Taq*I clones were selected. Bacterial cultures of the chosen clones were rearrayed from the original libraries to the 384-well microtiter plate filled with 50 μl of the "freezing medium" using a MicroGridII arrayer (Biorobotics). After rearraying, the cultures were allowed to grow for 24 h at 37°C. A 0.5 μl aliquot of each culture was used as a template for insert amplification in the conditions as described in "Preparation of DArT arrays". The amplified inserts (markers) were sequenced in the Biomolecular Resource Facility (BRF) of the Australian Cancer Research Foundation (ACRF), Australian National University, Canberra, Australia. The sequencing reactions were carried out in-house according to the methods recommended by BRF (www.brf.jcs.anu.edu.au/services/DNAsequencing). Amplified PCR products were agarose-gel purified and sequenced on an ABI 3730 sequencer (ACRF BRF, ANU) following the manufacturer's protocol (Applied Biosystems 2002). The traces were viewed in Sequence Scanner software v. 1.0 (Applied Biosystems). The 31 sequences of either poor quality (the basecaller quality value less than 20 for majority of chromatogram) or too short (less than 50 bp) were rejected. A Perl script developed in-house (Diversity Arrays Technology Pty. Ltd, unpublished) was used to trim the adaptor sequences, or up to 30 bp upstream of *Eco*RI cloning site from both the 3′ and 5′ ends if adaptor sequences could not be read. Trimmed sequences were subjected to NBLAST and TBALSTX searches against the nonredundant (nr) and the EST section of GenBank with an E value threshold of 1e-04. The searches were done in batches using a custom-written Perl script based on NCBI CGI BLAST interface (Diversity Arrays Technology Pty. Ltd, unpublished). The sequences were also analyzed for redundancy by using pairwise comparisons to detect sequence homology over a minimum of 100 bp. This was performed using stand alone BLAST v. 2.2 (NCBI) as well as scripts written in Perl and Mathcad v. 11.2a (Mathsoft) (D. Jaccoud, personal communication).

Results

Evaluation of the methods for genome complexity reduction

The method of genome complexity reduction is the first step in DArT technology. Resulting genomic representations are the starting point for the downstream steps in DArT genotyping. A number of approaches to reduce the genome complexity can be utilized. We based our methods on either solely restriction-enzyme (RE) digestion of DNA, adaptor ligation and amplification of adaptor-ligated fragments (methods 1–8 and method 9, Online Resource 1a and 1b, respectively) or a combination of this approach with exploiting naturally occurring Miniature Inverted-repeat Transposable Elements (MITEs) in plants (method 10, Online Resource 1c). In summary, we used various combinations of single or double primary cutters (6 bp) with one to five secondary cutters (4 bp) followed by the ligation of one to five adaptors and amplification of the adaptors' ligated fragments. Each method was tested on a set of 16 sugarcane genotypes representing old cultivars present in the genealogy of most cultivars worldwide, important cultivars or elite parents in Australia and

Table 2 Performances of the tested methods of genome complexity reduction

Method number	Complexity reduction method = library = array	Number of analyzed clones	Percentage of polymorphic clones	Average P^a	Average call rate[b]
1	***Pst*I/*Taq*I**	**1,536**	**7.03**	**80.68**	**92.54**
2	***Pst*I/*Mse*I**	**1,536**	**6.64**	**80.53**	**92.94**
3	*Afl*III/*Taq*I/*Alu*I	768	1.43	81.81	85.46
4	*Pst*I/*Eco*RI/*Bst*NI/*Taq*I	1,536	1.37	81.20	91.78
5	*Pst*I/*Eco*RI/*Bst*NI	768	3.25	79.78	90.67
6	*Pst*I/*Eco*RI/*Taq*I	768	2.21	81.66	92.67
7	*Pst*I/*Hpa*II/*Dpn*II	1,536	0.52	80.82	89.69
8	*Pst*I/*Hpa*II/*Taq*I	1,536	5.33	79.59	88.69
9	*Pst*I/*Taq*I recut with *Dpn*II/*HpyCH4*IV/ *Mse*I/*Nla*III	768	3.38	79.67	86.83
10	**MITE/*Bsp1286*I**	**3,072**	**9.78**	**81.36**	**92.51**

The best genome complexity reduction methods are shown in boldface

[a] Function measuring the level of relative target signal bimodality/polymorphism for a particular spot

[b] The percentage of successful allelic assignments

including two genotypes of each of the ancestral species *S. officinarum* and *S. spontaneum* (Table 1). The results of polymorphism analysis for each of the tested methods are summarized in Table 2. The polymorphism frequencies varied almost 20-fold between the methods, in the range from 0.52% (method 7) to 9.78% (method 10) with an average of 4.09% ($P < 0.0001$ as determined by a Pearson's χ^2 statistic). Based on the results presented in Table 2 and criteria described in "Methods" in the section "Evaluation of the genome complexity reduction methods", three methods of genome complexity reduction, *Pst*I/*Taq*I, *Pst*I/*Mse*I, and MITE/*Bsp1286*I, were the most promising with 7.03, 6.64, and 9.78% polymorphism frequency, respectively (Table 2; methods 1, 2 and 10).

In order to validate the performance of the *Pst*I/*Taq*I method in reducing the genome complexity, genomic DNAs of one *S. officinarum*, one *S. spontaneum*, and one modern cultivar were hybridized to the *Pst*I/*Taq*I array without reducing their genome complexity. For each sample tested, approximately only 5% of the genomic DNA hybridized to this array (Fig. 1, panels a and c), indicating the effectiveness of *Pst*I/*Taq*I clones in simplifying the genome complexity. We estimate that the *Pst*I/*Taq*I genomic representation has in excess of 10,000 individual fragments representing 0.05–0.1% of the genome and

reduces its complexity by filtering away the repetitive sequences (see "Discussion" for the details).

We tested the *Pst*I/*Taq*I method's reproducibility and robustness by analyzing two sugarcane accessions (*S. officinarum* Q813 and cultivar F172), as described in "Methods", on the array containing 1,536 clones printed in triplicate. Analysis of the 24 arrays (12 for each genotype) for a relative signal intensity at each probe on the array indicated that DArT assays based on *Pst*I/*Taq*I complexity reduction method are highly reproducible with an R^2 of approximately 0.98 (Online Resource 3). In addition, this experiment identified 55 DArT markers among 1,536 probes (3.6%), discriminating Q813 and F172 with 100% consistency (data not presented).

Enrichment for the polymorphic markers

We attempted Suppression Subtractive Hybridization (SSH) (Diatchenko et al. 1996) to enrich for the polymorphic markers (see "Discussion" for rationale). The results of the SSH experiment are shown in Table 3. The average polymorphism frequency of all subtraction variants (one or two rounds of subtraction, 30:1 or 10:1 quantitative ratio of driver to tester—method 10 A, B, C, D in Table 3) was only slightly better than the *Pst*I/*Taq*I method only, 7.63 and 7.03%, respectively (Tables 2, 3, method 1,

Fig. 1 Effectiveness of the genomic representations based on the *Pst*I/*Taq*I method in reducing the genome complexity of sugarcane. The *Pst*I/*Taq*I based representations effectively filter away the repetitive sequences: ~95% of the *Pst*I/*Taq*I clones contain low copy sequences. **a** A *Pst*I/*Taq*I array hybridized with genomic DNA of sugarcane cultivar Q124 labelled with Cy3 dye; **b** same array hybridized with the reference DNA labelled with FAM dye; **c** a graph representing median signal intensity of genomic DNA of three sugarcane genotypes hybridized to the same array
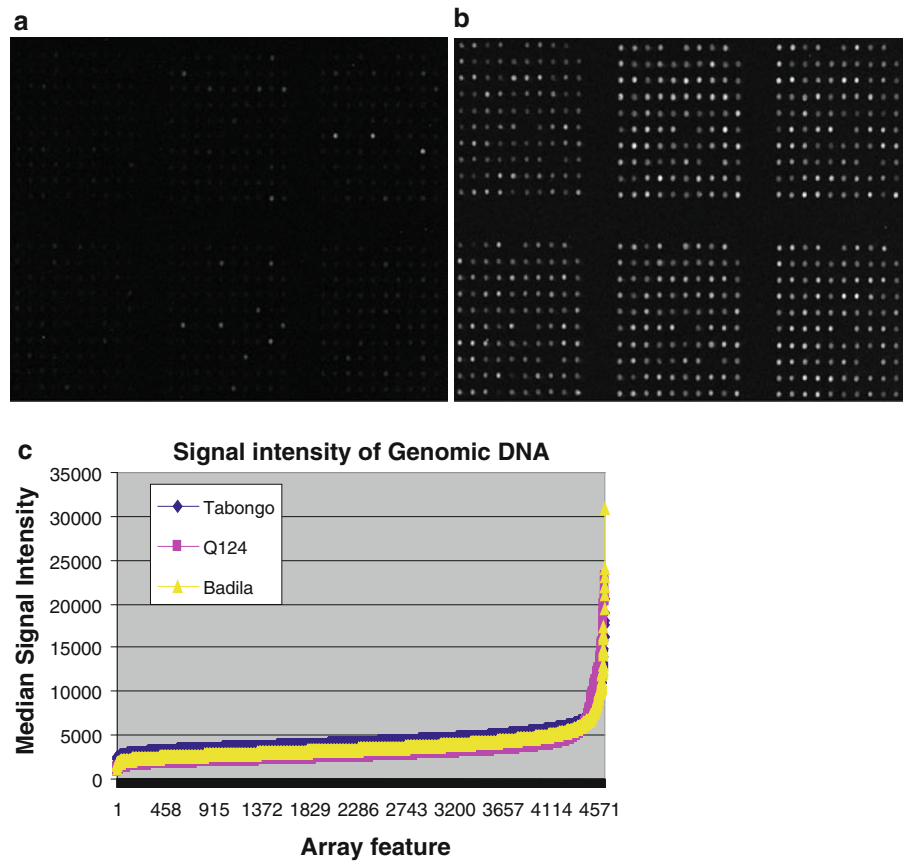

Table 3 Summaries of the results of suppression subtractive hybridization experiment to enrich for polymorphic clones

Number of analyzed clones		Percentage of polymorphic clones	Average P^{a}	Average call rate[b]
A	1,152	13.28	83.27	93.28
B	1,152	7.12	82.09	92.81
C	1,152	6.51	81.76	92.81
D	1,152	3.64	80.30	90.88
Sum/average A–D	4,608	7.63	81.55	92.44

Prior to SSH technique, the genome complexity was reduced with the method *Pst*I/*Taq*I recut with *Dpn*II/*Hpy*CH4IV/*Mse*I/*Nla*III

A—one round subtraction, 30:1 ratio of driver to tester, B—one round subtraction, 10:1 ratio of driver to tester, C—two rounds subtraction, 30:1 ratio of driver to tester, D—two rounds subtraction, 10:1 ratio of driver to tester

[a] Function measuring the level of relative target signal bimodality/polymorphism for a particular spot

[b] The percentage of successful allelic assignments (for experiment details see section "An approach to enrich for polymorphic clones" in "Methods")

respectively). However, one-round subtraction with a 30:1 quantitative ratio of driver to tester resulted in 13.28% polymorphism frequency, almost doubling the result obtained for the *Pst*I/*Taq*I method only ($P < 0.0001$ based on a Pearson's χ^2 statistic). To

compare these two methods directly, another experiment was performed on the expanded *Pst*I/*Taq*I array (3,456 clones) with 2,688 clones of a *Pst*I/*Taq*I/SSH array (method 10 A, B, C, Table 3) by hybridization of genomic representations from two genotypes of
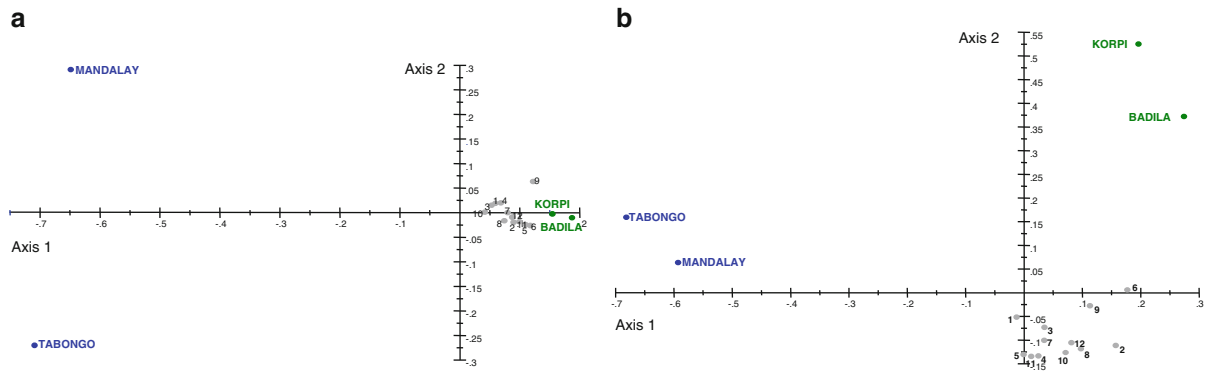
a

b

Fig. 2 Comparison of the performance of the *Pst*I/*Taq*I DArT markers with and without subtraction (*SSH*) in revealing the genetic relationships in sugarcane. The *Pst*I/*Taq*I SSH markers perform better in determining the pattern of genetic relationships among sugarcane genotypes representative of *Saccharum* diversity, as revealed by two-dimensional plots of factorial analysis (*FAC*) based on a Jaccard dissimilarity matrix. FAC was computed from presence/absence of 125 *Pst*I/*Taq*I (**a**) and 103 SSH DArT markers (**b**) for four ancestral species Mandalay and Tabongo (*S. spontaneum*), Korpi and Badila (*S. officinarum*), and twelve cultivars: *1*—F172, *2*—QN58-829, *3*—CP74-2005, *4*—Tellus, *5*—Q201, *6*—Trojan, *7*—Mida, *8*—CO475, *9*—Trojan, *10*—QN67-3184, *11*—Q124, *12*—Q208

each ancestral species *S. spontaneum* and *S. officinarum,* and 12 cultivars (Table 1). Each representation was hybridized in four independent replications. The polymorphism frequencies obtained were similar to the previous results with 7.71 and 6.34%, respectively for the *Pst*I/*Taq*I/SSH-based array and the *Pst*I/*Taq*I-only array. However, because sugarcane ancestral species, especially the "wild" species *S. spontaneum*, are strongly differentiated from modern cultivars on the genetic level (Grivet and Arruda 2001), they contribute the most to the detectable polymorphism. Accordingly, once four ancestral species were removed from the polymorphism analysis, the *Pst*I/*Taq*I/SSH-based array was better in identifying polymorphism among modern cultivars with 6.32% polymorphism frequencies compared to 3.76% for *Pst*I/*Taq*I array ($P < 0.0001$ based on a Pearson's χ^2 statistic).

To test the performance of the *Pst*I/*Taq*I versus *Pst*I/*Taq*I/SSH polymorphic markers, the markers generated in the described experiment were used in a Factorial Analysis of Correspondence (FAC). The analysis was based on a Jaccard dissimilarity matrix computed from the presence/absence of alleles at 125 and 103 markers, respectively, for *Pst*I/*Taq*I and *Pst*I/*Taq*I/SSH, using the DARwin5 software (Perrier and Jacquemoud-Collet 2006). Figure 2 illustrates the improved performance of *Pst*I/*Taq*I DArT markers subjected to SSH in better determination of a genetic relationship pattern within the population of analyzed

sugarcanes, especially in separation of *S. officinarum* ancestral species from the modern cultivars. As expected, both *Pst*I/*Taq*I and *Pst*I/*Taq*I/SSH markers revealed the "wild" ancestral species *S. spontaneum* (Mandalay and Tabongo) as the most distant. For *Pst*I/*Taq*I markers (Fig. 2a), Axis 1 (73.5% of total variance) separated *S. spontaneum* (Mandalay and Tabongo) from *S. officinarum* (Korpi and Badila), which grouped together with the cultivars (11.3% of variation detected, Axis 2). Once a part of the common component of the *S. officinarum* genome was removed by subtraction, the genetic differences between *S. spontaneum, S. officinarum*, and the cultivars were much better resolved (Fig. 2b). The separation between *S. spontaneum* and other genotypes was along Axis 1 (with 41.7% of total variance explained) while *S. officinarum* separated from the modern cultivars along Axis 2 (23.2% of total variance).

Genetic relationships among sugarcane accessions revealed by DArT markers

The DArT markers of the two best complexity reduction methods MITE/*Bsp1286*I and *Pst*I/*Taq*I/SSH (method 10 in Tables 2, 3, respectively) were tested for their performance in revealing the genetic relationship between 16 sugarcane genotypes used for researching the genome complexity reduction methods and building the arrays (Table 1). The
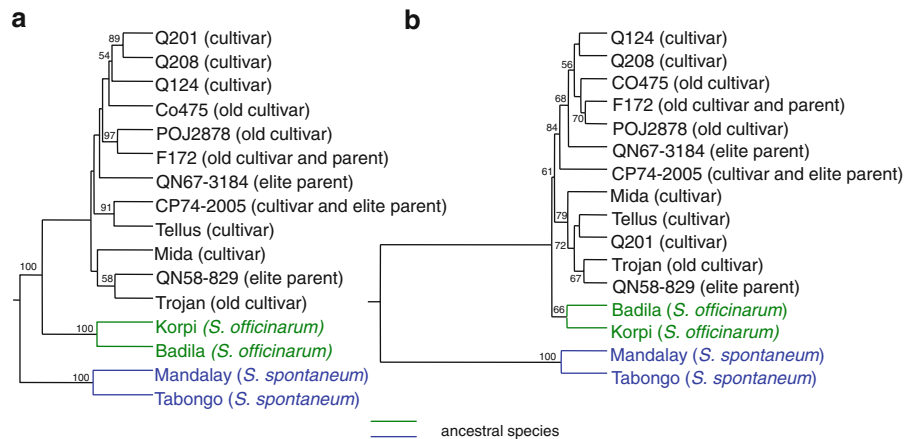
Fig. 3 Genetic relationships between sugarcane genotypes representative of *Saccharum* diversity revealed by the DArT markers. UPGMA dendrograms based on the Nei/Li restriction-fragment-distance index for four ancestral sugarcane species and twelve cultivars based on 364 polymorphic markers obtained with the *Pst*I/*Taq*I SSH genome complexity reduction method (**a**) and 303 polymorphic markers obtained with the MITE/*Bsp*1286I genome complexity reduction method (**b**). The *numbers* at the *branches* are confidence values based on Felsenstein's bootstrap method shown only if greater than 50% for 1,000 replicates

library created with the *Pst*I/*Taq*I/SSH method was first expanded to 6,144 clones and the genotyping experiment for 16 samples was repeated on this new array with three independent replications of each sample. A total of 364 markers were scored by DArTSoft with the default settings. These markers, along with 303 markers obtained for the same set of 16 genotypes with the MITE/*Bsp*1286I method, were used to construct two UPGMA dendrograms based on the Nei/Li restriction-fragment-distance index using the PHYLIP 3.6 software package (Felsenstein 2002). As shown in Fig. 3, the relationships revealed among the assayed sugarcane genotypes appear consistent with expectations and results obtained with other technologies: the ancestral species of *S. spontaneum* and *S. officinarum* separated from the rest of the samples with both methods of genome complexity reduction. As expected, *S. spontaneum* samples Tabongo and Mandalay were the most distant and formed individual clusters on the dendrogram. The *S. officinarum* Korpi and Badila, although separated from modern clones and cultivars on a separate branch, still formed one clade with them, in agreement with the notion that ~80% of the modern sugarcane genome is derived from *S. officinarum* species (D'Hont et al. 1996). The consistency of genetic relationship revealed by the complexity reduction methods surveyed is noteworthy since they were targeting different types of DNA sequences and

use different enzymatic methods. However, maximum bootstrap values (1,000 replicates) obtained for the *Pst*I/*Taq*I/SSH method, i.e. 100% for both ancestral species (Fig. 3a) versus 100% for *S. spontaneum* but only 66% for *S. officinarum* in the MITE/*Bsp*1286I method (Fig. 3b), showed that the *Pst*I/*Taq*I/SSH method provided a higher level of confidence in defining the genetic relationship. The UPGMA dendrograms supported the result obtained already in FAC for the *Pst*I/*Taq*I/SSH versus the *Pst*I/*Taq*I method (Fig. 2).

Genetic behavior of DArT markers in sugarcane

The segregation of DArT markers in sugarcane was assessed on 94 progeny genotypes from the cross IJ76-514 (*S. officinarum*) × Q165 (an Australian cultivar). The genotyping experiment was carried out on the array consisting of 6,144 clones (Online Resource 2). Individual progenies (in duplicate) were hybridized to this array and DArTSoft analysis with the default settings resulted in 241 polymorphic markers giving 3.9% polymorphism frequency. In the absence of segregation distortion, the expected segregation ratios for the tested mapping population were 1:1 for single-dose and 3:1 for both bi-parental single-dose markers and double-dose markers present only in one parent. The majority of the polymorphic markers (204, 84.7%) segregated as single-dose

Table 4 Summary of the segregation patterns of 241 polymorphic DArT markers scored for the IJ76-514 × Q165 mapping population

	Number	Percentage
Polymorphic markers	241	100
Polymorphic markers between parents	215	89.21
Number of segregating markers	204	84.65
Single-dose markers (1:1)	149	73.04
Bi-parental single-dose markers or double-dose markers present in one parent only (3:1)	55	26.96
Markers with distorted segregation	37	15.35

markers or double-dose markers with either 1:1 or 3:1 ratios. Of these 204 markers, 73% segregated as single-dose markers, while 27% segregated as bi-parental single-dose markers (or double-dose

markers present only in one parent). The parents scored differentially for 215 markers (89.2%) with the majority of scoring assignments "1" coming from the Q165 (an Australian cultivar) parent. The IJ76-514 (*S. officinarum*) parent had only 21% of markers scored representing "1" allele. The segregation patterns of the polymorphic markers are summarized in Table 4, while Fig. 4 shows an example of DArT markers incorporation into linkage groups of homo (eo)logy group II of a framework genetic map of the Q165 sugarcane cultivar (Aitken et al. 2005).

Analysis of sugarcane candidate polymorphic DArT markers' sequences

We sequenced 384 candidate polymorphic markers chosen as described in "Methods" and analyzed 353
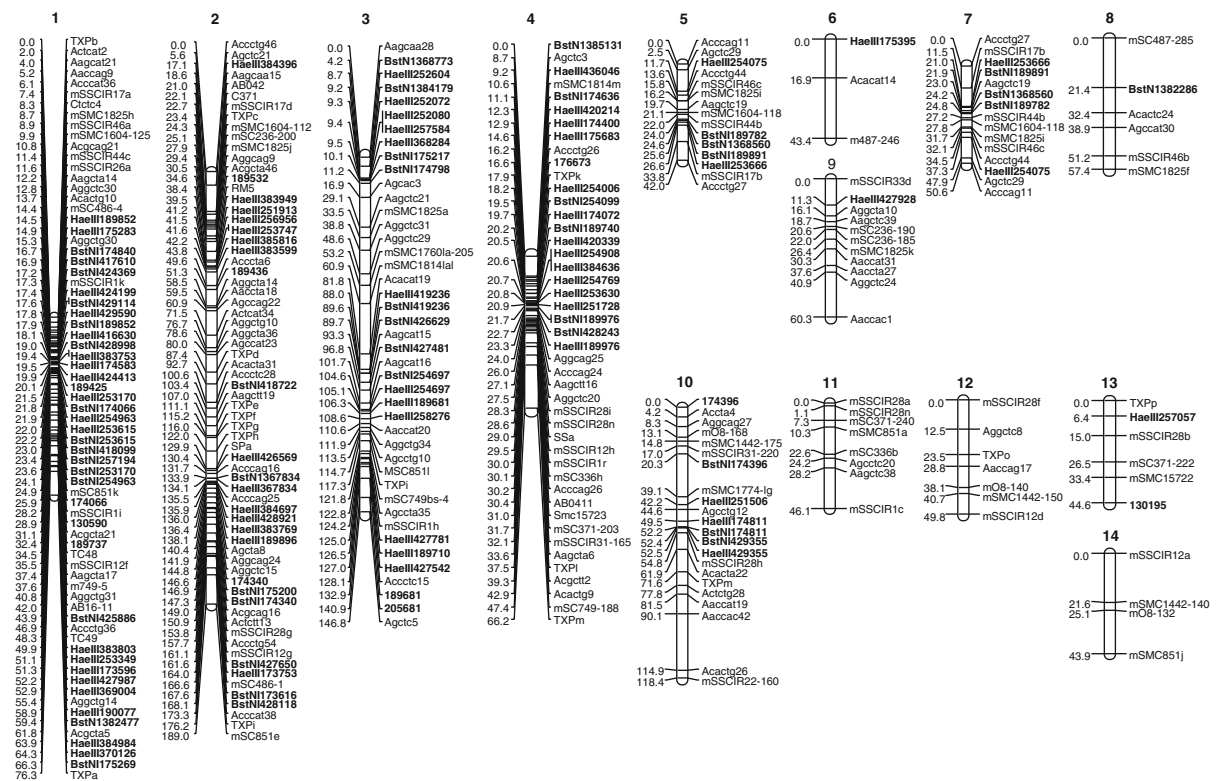

Fig. 4 Location of DArT markers on linkage groups from homo(eo)logy group II on a genetic linkage map of the Q165 sugarcane cultivar. The DArT markers were incorporated into the Q165 linkage map generated by Aitken et al. (2005) using Joinmap v 3.0 as in Aitken et al. (2007). The DArT markers are shown in *boldface* and annotated according to the method of genome complexity reduction used to generate them: *BstN*I followed by the *6-digit number* represents the *Pst*I/*Taq*I/*BstN*I method, *Hae*III followed by the *6-digit number* represents the *Pst*I/*Taq*I/*BstN*I method, and the *6-digit number* represents the *Pst*I/*Taq*I method. The annotations of the other types of markers are as in Aitken et al. (2005)

Table 5 Results of BLAST searches performed for the sugarcane DArT markers against NCBI nucleotide databases

	NBLAST	TBLASTX
Number (%) similarities found in non-redundant database	89 (25.2%)	102 (28.9%)
Number (%) similarities found in EST database[a]	147 (41.6%)	153 (43.3%)
Number (%) similarities found in *S. officinarum* EST sequences	102 (28.9%)	109 (30.9%)

[a] Non-mouse and non-human EST entries

high-quality sugarcane sequences with a phred quality value greater than or equal to 20, average sequence length of 314 bp, minimum length of 57 bp and maximum length of 797 bp. These sequences were subjected to searches in NCBI nucleotide databases. NBLAST and TBALSTX searches against the nonredundant (nr) and the EST section of GenBank were performed. The results of the searches are summarized in Table 5. The search in the nucleotide nonredundant database returned 25 and 29% similarities found for untranslated and translated sequences, respectively. This proportion went up to more than 40% in both cases, when the search was narrowed to the non-mouse and non-human EST database. It is worth pointing out that approximately 29 and 31% similarities for untranslated and translated searches, respectively, were found for sugarcane EST entries.

The detailed analysis of sequenced sugarcane DArT markers is in progress, though a brief glimpse of initial results revealed some interesting similarities. For example, for one of the clones a very strong similarity ($10^{-39} \leq E \leq 10^{-70}$) was found (nr, NBLAST, TBALSTX) with receptor-like kinase (*rlk*) gene of other Poaceae such as maize, rice, oat, barley and wheat; in the last species the *rlk* gene is implicated in leaf rust resistance (Feuillet et al. 1997). Strong similarities were also identified with sequences of wild relatives of modern sugarcane cultivars from the so-called "*Saccharum* complex": *Erianthus arundinaceus* Retz. and several representatives of *Miscanthus* genus. For example, a strong similarity was found to the 18S rRNA gene and a ribosomal DNA external transcribed spacer of *E. arundinaceus* ($E \leq 10^{-43}$ NBLAST, $E \leq 10^{-48}$ TBLASTX) and to the 25S rRNA and 17S rRNA genes and an internal transcribed spacer 1 of *Miscanthus sinensis*

Anderss. ($10^{-8} \leq E \leq 10^{-10}$ NBLAST, $10^{-10} \leq E \leq 10^{-15}$ TBLASTX).

The redundancy of sequenced markers was analyzed by sequence pairwise comparison. In total, there were 36 redundant markers: one present in five copies, five in three copies and 22 in two copies. Twenty-seven (75%) redundant markers originated from *Pst*I/*Taq*I/SSH libraries generated from IJ76-514 and Q165, the parents of the mapping population (*Pst*I/*Taq*I/SSH mapping); six redundant clones came from *Pst*I/*Taq*I/SSH libraries constructed from 16 sugarcane genotypes representing sugarcane diversity (*Pst*I/*Taq*I/SSH diversity), and only three were from the *Pst*I/*Taq*I library constructed from 16 sugarcane genotypes representing the *Saccharum* diversity (*Pst*I/*Taq*I diversity). Taking into account the origin of sequenced clones resulted in 15.6, 4.6 and 6.0% redundancy for *Pst*I/*Taq*I/SSH mapping, *Pst*I/*Taq*I/SSH diversity and *Pst*I/*Taq*I diversity libraries, respectively, and 10.2% on average.

Discussion

Diversity Arrays Technology genotyping is a well-established technology for a number of species providing hundreds to thousands of high-quality polymorphic markers. The technology finds an increasingly broad range of applications, for example in whole-genome-based molecular breeding, genetic identity testing, diversity analysis, and molecular systematics of natural populations (www.Diversity Arrays.com/publications). We report here on a development of DArT technology for sugarcane. The most critical step in DArT is the method of genome complexity reduction, as it substantially determines the efficiency of polymorphism identification. Capturing the polymorphic fraction of the genome is especially challenging in the polyploid species, where copies of homologous chromosomes "dilute" the polymorphism. Therefore, our primary effort in developing DArT for sugarcane relied on an extensive testing of various methods to reduce genome complexity of a pool of sugarcane accessions representing the *Saccharum* genetic diversity.

The successful performance of DArT in identifying polymorphism relies largely on high-fidelity restriction enzymes (RE) detecting SNPs in RE recognition sequences (Akbari et al. 2006; Wittenberg et al. 2005).

Therefore we evaluated 12 restriction enzymes in 10 various methods for their effectiveness in polymorphism identification. As shown in the "Results" section, the methods based on *Pst*I restriction enzyme as a primary cutter performed well in finding polymorphism in sugarcane, as was the case in other plant species for which DArT genotyping was developed. For example, the *Pst*I/*Taq*I method resulted in 10.4% polymorphism in barley (Wenzl et al. 2004), 14.6% in cassava (Xia et al. 2005), and 9.4% in wheat (Akbari et al. 2006). The good performance of *Pst*I RE can be attributed to the methylation susceptibility of the enzyme. Its recognition sequence contains two CXG groups (X = A or T) and methylation of cytosine blocks the enzyme activity. Since the repetitive DNA blocks in plant genomes are usually methylated at 5′-CG-3′ and 5′-CXG-3′ cytosine (Gruenbaum et al. 1981), *Pst*I preferentially cuts non-methylated, non-repetitive DNA. Thus, the *Pst*I-based complexity reduction methods filter away repetitive sequences and the markers obtained from such methods are predominantly located in low-copy, gene-rich areas of the genome which are typically hypomethylated or non-methylated at CG and CXG sites (Rabinowicz et al. 2003).

MITE/*Bsp1286*I was another effective method of reducing genome complexity and identifying polymorphism in sugarcane. This method was also successfully used for DArT genotyping in rice and sorghum (S. Patarapuwadol, personal communication; unpublished data). It is not surprising as MITEs are widespread in plants, small in size but high in copy number and polymorphism, stable and located preferentially in the genic regions of the genomes. These features offer a good basis from which to exploit MITEs as molecular markers and indeed they are becoming popular as useful tools in plant genetics and genomics (Casa et al. 2000; Park et al. 2003; reviewed in Feng 2003). An application of *Bsp1286*I restriction enzyme increased further the MITE-identified polymorphism, since the enzyme recognition site 5′-GDGCHC-3′ (D = A or G or T; H = A or C or T) allows for a variety of sequence combinations. Three of these combinations were captured by the design of adaptors annealing to the products of *Bsp1286*I digestion.

With polymorphism frequencies of 7.03 and 9.78% for the two best methods of genome complexity reduction *Pst*I/*Taq*I and MITE/*Bsp1286*I, respectively, we considered a polymorphism enrichment technique to facilitate development of a final array comprising 5,000–7,000 polymorphic markers. As ~ 70 to 80% of the genome of modern sugarcane cultivars originates from *S. officinarum* with eight copies of the basic chromosome number ($x = 10$, $2n = 8x = 80$) (D'Hont et al. 1996), the approach was to remove this repetitive part of the genome by means of subtraction, thus revealing more of the genome polymorphism originating from *S. spontaneum*. We decided on Suppression Subtractive Hybridization (SSH), a PCR-based DNA subtraction method originally developed to generate differentially regulated or tissue-specific cDNA libraries (Diatchenko et al. 1996). We used the same principle, which was to equalize the quantities of DNA fragments by hybridization kinetics and subsequent selective amplification of target DNA fragments. Although the average polymorphism frequency for all subtraction variants tested was only slightly better than for the *Pst*I/*Taq*I method without subtraction, twofold enrichment in the number of polymorphic markers was achieved in one of the subtraction variants. Thus, SSH seemed to be a promising technique for polymorphism enrichment in sugarcane. However, a redundancy of markers has to be taken into consideration before committing to this technique as a method of choice for building the final array (discussed below).

The subtraction showed an additional advantage apart from improving the polymorphism frequency. The subtracted *Pst*I/*Taq*I clones were better than *Pst*I/*Taq*I clones without subtraction in determining the genetic relationships between surveyed sugarcane accessions as revealed by Factorial Analysis of Correspondence. This result was further confirmed when the *Pst*I/*Taq*I/SSH markers obtained in the independent experiment, as well as the MITE/*Bsp1286*I markers from the second best method of genome complexity reduction, were used to draw the UPGMA dendrograms. The genetic relationships revealed among sixteen genotypes based on these two sets of markers are consistent with the previous findings for sugarcane genetic relationships obtained with other molecular markers and pedigree information. Assessments of genetic diversity and similarity of various numbers of sugarcane accessions were performed with molecular markers such as rDNAs (Glaszmann et al. 1990), RFLPs (Jannoo et al. 1999b) and AFLPs (Besse et al. 1998; Lima et al. 2002;

Aitken et al. 2006). The findings of these studies demonstrated a strong genetic differentiation between *S. officinarum* and *S. spontaneum* and both ancestral species from the modern cultivars, just as was shown with both types of the best sugarcane DArT markers.

The segregation ratios of sugarcane DArT markers demonstrated that, in spite of high polyploidy and the bi-specific nature of the sugarcane genome, the majority of DArT markers segregated in the Mendelian fashion. The estimated proportion of single-dose markers obtained for the IJ76-514 × Q165 mapping population in this study is similar to the proportions reported for the same, as well as other, sugarcane mapping populations. There were 71% of polymorphisms detected with a combination of AFLPs, SSRs and RAFs markers presenting as single-dose markers in the IJ76-514 × Q165 mapping population (Aitken et al. 2005), while 73 and 79% were estimated, respectively, for a mapping population derived from a cross between two pre-commercial cultivars, using a combination of RFLPs, SSRs and AFLPs (Garcia et al. 2006) and the selfed progeny of R570 cultivar with AFLP markers (Hoarau et al. 2001).

From the markers scored differently between the parents, the majority of scoring assignments "1" originated from modern cultivar Q165 parent. The bi-specific nature of modern sugarcane cultivars may be the contributing factor to this finding. The higher redundancy levels of the *S. officinarum* part of the genome (Aitken et al. 2005) and known lower diversity of this species (Aitken et al. 2006) could have contributed to the under-representation of *S. officinarum* polymorphic markers. Thus, the majority of markers identified from Q165 seemed to originate from the *S. spontaneum* part of the genome due to its known more polymorphic nature. Better coverage of *S. spontaneum* than of *S. officinarum* chromosomes is a typical feature of sugarcane mapping (Grivet and Arruda 2001).

In order to obtain extensive coverage of each chromosome for a species with such a large and complex genome, higher numbers of single-dose markers would be required. The dominance of single-dose markers is a desirable characteristic in out-crossing polyploid species such as sugarcane, since alleles present as a single dose are much more informative in the construction of genetic maps and marker-assisted selection (MAS) than any others. Single-dose markers in the coupling phase are used to construct the map and in the repulsion phase to determine chromosome assortment. Duplex or multiplex markers can only be incorporated into the framework map constructed based on single-dose markers. Therefore, experiments aiming to enrich for single-dose sugarcane DArT markers are well advanced. Incorporating the third frequently cutting restriction enzyme *Bst*NI or *Hae*III into the *Pst*I/*Taq*I method of genome complexity reduction resulted in a 25% increase in the frequency of single-dose markers (data not presented). While the detailed genetic map of the Q165 cultivar comprising the single-dose DArT markers resulting from the described enrichment will be published separately, Fig. 4 shows an example of incorporation of these DArT markers into the Q165 linkage map generated by Aitken et al. (2005). The majority of the markers mapped into the existing linkage groups, indicating that they map to similar regions of the genome as AFLP and SSR markers.

The sequence analysis of sugarcane DArT markers revealed some interesting findings. Close to one-third of these markers were found to be similar to sugarcane EST sequences originating from the transcribed portion of the genome. As these markers are derived from expressed genes, they may facilitate functional genomics studies and identification of the genetic basis of desirable traits (Varshney et al. 2005). It is especially of interest in sugarcane, the species where until 2006 only three major genes were identified: two rust resistance genes and a gene controlling stalk color (Asnaghi et al. 2004; Raboin et al. 2006). The similarities found for some DArT markers and entries for *Erianthus* sp. and *Miscanthus* sp. also deserve highlighting. Both plants are related genera within the Sacchararinea tribe, though it was recently suggested that lineages leading to the genus *Saccharum* were independent from those leading to genera *Erianthus* and *Miscanthus* (Grivet et al. 2004). However, there is considerable interest in incorporating *E. arundinaceus* Retz., characterized by high fiber content, vigor, and resistance to disease and environmental stresses, into sugarcane breeding programs, as fertile hybrids of *E. arundinaceus* and sugarcane have been established (Cai et al. 2005). Thus DArT markers bearing this similarity could be of use in assessing the phylogenetic relationship of sugarcane and related species and in facilitating marker-assisted incorporation of parts of the *E. arundinaceus* genome into modern sugarcane cultivars.

There was 10.2% redundancy observed for sequenced sugarcane markers contributed mainly (75%) by clones originating from two subtracted parents of the mapping population. This finding may indicate that the SSH-mediated polymorphism enrichment would be better suited for the more diverse genetic pool. Indeed, the subtracted clones coming from 16 sugarcane genotypes contributed only 17% of redundant markers. Overall, the redundancy level found for sugarcane was significantly lower than the redundancy noted for barley and *Arabidopsis thaliana* L. DArT markers (Wenzl et al. 2004; Wittenberg et al. 2005) and almost identical to that recorded for wheat DArT markers (Akbari et al. 2006). The redundancy estimation in barley and wheat was carried out on the mapping data based on identity of marker scores, which could lead to an overestimation since the closely linked loci would score the same but could still be different sequences. The more accurate redundancy assessment based on the sequence analysis carried out for *A. thaliana* revealed approximately 1.78-fold redundancy (42%) (Wittenberg et al. 2005). The difference observed for *Arabidopsis* and sugarcane could be attributed to the extreme differences in these species genome sizes, ploidy levels and the use of a single accession for construction of the *A. thaliana* array.

Sequencing of all DArT markers discovered (ongoing) will offer not only an ability to remove the excessive redundancy from the final genotyping array, but will also facilitate relating the results of sugarcane genetic analysis to the genome sequence resources of sorghum. The sorghum genome sequencing project has been completed (Paterson et al. 2009), and therefore the resulting sequence assembly will provide a useful framework for sorghum and sugarcane genetics, as the two crops have highly related genomes (Guimarães et al. 1997). A recent development of DArT for sorghum (Mace et al. 2008) and sequencing of over 2,000 sorghum markers (Bouchet, personal communication) will further facilitate the transfer of genetic and genomic information between these two important crops. However, any strongly reductionist attempts to "translate" sorghum genome information into genetics of sugarcane are likely to fail, as the two organisms vary greatly in their ploidy levels, genome stability and level of DNA sequence variation within genotypes and breeding populations. Indeed, each sugarcane cultivar could be viewed genetically as a "bulk" with many alleles representing a substantial fraction of total diversity segregating in breeding populations. In this context it is worth noting that the DArT platform has been deployed successfully for quantitative Bulked Segregant Analysis (qBSA) in barley and that the relative signal intensity of the representations prepared from bulks was shown to monitor precisely allele frequencies in complex samples (Wenzl et al. 2007). It is our view that highly reproducible sugarcane DArT assays open the possibility of using relative signal intensity of DArT arrays (in addition to binary scored markers) in allele frequency estimation and effective marker–trait association analysis in sugarcane. We will present results in support of this notion in our next paper.

Conclusions

The presented work demonstrates that DArT genotyping effectively discovers and scores hundreds of polymorphisms in the large, aneupolyploid genome of sugarcane. Sugarcane DArT fingerprints can be used for both investigation of genetic relationships and construction of genetic maps. An expansion of libraries generated with the best genome complexity reduction methods will enable us to construct a "sugarcane DNA chip" capable of providing genome-wide scans with many thousands of markers for hundreds of genotypes timely and cost-effectively. This could open up a gateway for whole-genome—profile-assisted breeding programs of sugarcane, providing superior varieties in a much shorter time than via the currently used conventional breeding methods. The methods of genome complexity reduction utilizing the restriction enzymes *Pst*I/*Taq*I and incorporating the third frequently cutting RE either *Bst*NI or *Hae*III, due to their best performance, became our method of choice to generate genomic representations for routine sugarcane genotyping.

This new capability of utilizing DArT-based whole genome information in practical breeding is in fact already at the early stages of implementation by the Australian sugarcane breeding program. While it will take a few years before the anticipated advantages of genome-based selections can be validated, our modeling based on large volume of phenotypic and DArT marker data suggests that

sugarcane is well positioned to capture the value of genome profiling technology presented in this paper.

References

Aitken KS, Jackson PA, McIntyre CL (2005) A combination of AFLP and SSR provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. Theor Appl Genet 110:789–801

Aitken KS, Jackson PA, Piperidis G, McIntyre CL (2006) AFLP analysis of genetic diversity within S. officinarum and comparison with sugarcane cultivars. Aust J Agric Res 57:1167–1184

Aitken KS, Jackson PA, McIntyre CL (2007) Construction of a genetic map for Saccharum officinarum incorporating both simplex and duplex markers to increase genome coverage. Genome 50:742–756

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathmell B, Huttner E, Kilian A (2006) Diversity Arrays Technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet 113:1409–1420

Asnaghi C, Roques D, Ruffel S, Kave C, Hoarau JY, Telismart H, Girard JC, Raboin LM, Risterucci AM, Grivet L, D'Hont A (2004) Targeted mapping of a sugarcane rust resistance gene (Bru1) using bulked segregant analysis and AFLP markers. Theor Appl Genet 108:759–764

Besse P, Taylor G, Carroll B, Berding N, Burner D, McIntyre CL (1998) Assessing genetic diversity in a sugarcane germplasm collection using an automated AFLP analysis. Genetica 104:143–153

Biomolecular Resource Facility in JCSMR, ANU website [http://brf.jcs.anu.edu.au/services/DNAsequencing/]

Bonin A, Paris M, Despres L, Tetreau G, David J-P, Kilian A (2008) A MITE-based genotyping method to reveal hundreds of DNA polymorphisms in an animal genome after a few generations of artificial selection. BMC Genomics 9:459

Botha F (2006) Potential of sugarcane as biomass producer and biofactory [abstract]. Tropical Crop Biotechnology Conference Bulletin 8

Cai Q, Aitken K, Deng HH, Chen XW, Cheng F, Jackson PA, Fan YH, McIntyre CL (2005) Verification of intergeneric hybrids (F1) from Saccharum officinarum × Erianthus arundinaceus, and BC1 from F1 × sugarcane (Saccharum spp.) clones using molecular markers. Plant Breed 124:322–328

Casa A, Bronwer C, Nagel A, Wagel L, Zhang Q, Kresovich S, Wessler SR (2000) The MITE family Heartbreaker (Hbr): molecular markers in maize. Proc Natl Acad Sci USA 97:10083–10089

Cordeiro GM, Taylor GO, Henry RJ (1999) Characterization of microsatellite markers from sugarcane (Saccharum sp.), a highly polyploid species. Plant Sci 155:161–168

D'Hont A (2005) Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. Cytogenet Genome Res 109:27–33

D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glaszmann JC (1996) Characterization of the double genome structure of modern sugarcane cultivars (Saccharum spp.) by molecular cytogenetics. Mol Gen Genet 250:405–413

D'Hont A, Souza GM, Menossi M, Vincentz M, Van-Sluys AM, Glaszmann JC, Ulian E (2008) Sugarcane: a major source of sweetness, alcohol and bio-energy. In: Moore PH, Ming R (eds) Tropical crop plant genomics. Springer, New York

Da Silva JAG, Sorrells ME, Burnquist WL, Tanksley SD (1993) RFLP linkage map and genome analysis of Saccharum spontaneum. Genome 36:782–791

Diatchenko L, Lau YF, Cambpbell AP, Chenchik A, Moqadam F, Huang S, Lukyanow S, Lukyanow K, Gurskaya N, Sverdlov ED, Siebert PD (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci USA 93:6025–6030

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:1–15

Felsenstein J (2002) PHYLIP (Phylogeny Inference Package) version 3.6a3. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle

Feng Y (2003) Plant MITEs: useful tools for plant genetics and genomics. Genomics Proteomics Bioinformatics 1:90–99

Feuillet C, Schachermayr G, Keller B (1997) Molecular cloning of a new receptor-like kinase gene encoded at the Lr10 disease resistance locus of wheat. Plant J 11:45–52

Food and Agriculture Organization of the United Nations website [http://faostat.fao.org/default.aspx]

Garcia AAF, Kido EA, Meza AN, Souza HMB, Pinto LR, Pastina MM, Leite CS, Ulian EC, Figueira A, Souza AP (2006) Development of an integrated genetic map of a sugarcane (Saccharum spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. Theor Appl Genet 112:298–314

Glaszmann JC, Lu YC, Lanaud C (1990) Variation of nuclear ribosomal DNA in sugarcane. J Genet Breed 44:191–198

Goldemberg J (2007) Ethanol for sustainable energy future. Science 315:808–810

Grivet L, Arruda P (2001) Sugarcane genomics: depicting the complex genome of an important tropical crop. Curr Opin Plant Biol 5:122–127

Grivet L, D'Hont A, Roques D, Feldmann P, Lanaud C, Glaszmann JC (1996) RFLP mapping in cultivated sugarcane (Saccharum spp.): genome organization in a highly polyploid and aneuploidy interspecific hybrid. Genetics 142:987–1000

Grivet L, Daniels C, Glaszmann JC, D'Hont A (2004) A review of recent molecular genetics evidence for

sugarcane evolution and domestication. Ethnobot Res Appl 2:9–17

Gruenbaum Y, Nareh-Many T, Cedar H, Razin A (1981) Sequence specificity of methylation in higher plant DNA. Nature 292:860–862

Guimarães CT, Sills GR, Sobral BW (1997) Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. Proc Natl Acad Sci USA 94:14261–14266

Hoarau JY, Offman B, D'Hont A, Risterucci AM, Roques D, Glaszmann JC, Grivet L (2001) Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.) I. Genome mapping with AFLP markers. Theor Appl Genet 103: 84–97

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:e25

James KE, Schneider H, Ansell SW, Evers M, Robba L, Uszynski G, Pedersen N, Newton AE, Russell SJ, Vogel JC, Kilian A (2008) Diversity arrays technology (DArT) for pan-genomic evolutionary studies of non-model organisms. PLoS One 3:e1682

Jannoo N, Grivet L, Seguin M, Dookun A, D'Hont A, Glaszmann JC (1999a) Linkage disequilibrium among modern sugarcane cultivars. Theor Appl Genet 99:1053–1060

Jannoo N, Grivet L, Seguin M, Paulet F, Domaingue R, Rao PS, Dookun A, D'Hont A, Glaszmann JC (1999b) Molecular investigation of the genetic base of sugarcane cultivars. Theor Appl Genet 99:171–184

Jannoo N, Grivet L, David J, D'Hont A, Glaszmann JC (2004) Differential chromosome pairing affinities at meiosis in polyploidy sugarcane revealed by molecular markers. Heredity 93:460–467

Kilian A, Huttner E, Wenzl P, Jaccoud D, Carling J, Caig V, Evers M, Heller-Uszynska K, Cayla C, Patarapuwadol S, Xia L, Yang S, Thomson B (2005) The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. In: Proceedings of the international congress in the wake of the double helix: from the green revolution to the gene revolution, 27–31 May 2003, Bologna, Italy, Avenue Media, pp 443–461

Lezar S, Myburg AA, Berger DK, Wingfield MJ, Wingfield BD (2004) Development and assessment of microarray-based DNA fingerprinting in *Eucalyptus grandis*. Theor Appl Genet 109:1329–1336

Lima ML, Garcia AA, Oliveira KM, Matsuoka S, Arizono H, De Souza CL Jr, De Souza AP (2002) Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). Theor Appl Genet 104:30–38

Mace ES, Xia L, Jordan DR, Halloran K, Parh DK, Huttner E, Wenzl P, Kilian A (2008) DArT markers: diversity analyses and mapping in *Sorghum bicolor*. BMC Genomics 9:26

Mantovani P, Maccaferri M, Sanguineti MC, Tuberosa R, Catizone I, Wenzl P, Thomson B, Carling J, Huttner E, DeAmbrogio E, Kilian A (2008) An integrated DArT-SSR linkage map of durum wheat. Mol Breed 22:629–648

Mudge J, Andersen WR, Kehrer R, Fairbanks DJ (1996) A RAPD genetic map of *Saccharum officinarum*. Crop Sci 36:1362–1366

Official website of Diversity Arrays Technology P/L [http://www.diversityarrays.com/publications]

Park KC, Lee JK, Kim NH, Shin YB, Lee JH, Kim NS (2003) Genetic variation in Oryza species detected by MITE-AFLP. Genes Genet Syst 78:235–243

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Perrier X, Jacquemoud-Collet JP (2006) DARwin software [http://darwin.cirad.fr/france]

Rabinowicz PD, Palmer LE, May BP, Hemann MT, Lowe SW (2003) Genes and transposons are differentially methylated in plants but not in mammals. Genome Res 13: 2658–2664

Raboin LM, Oliveira KM, Lecunff L, Telismart H, Roques D, Butterfield M, Hoarau JY, D'Hont A (2006) Genetic mapping in sugarcane, a high polyploid, using bi-parental progeny: identification of a gene controlling stalk color and a new rust resistance gene. Theor Appl Genet 112:1382–1391

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. A laboratory manual. Cold Spring Harbour Laboratory Press

Sessittsch A, Hackl E, Wenzl P, Kilian A, Kostic T, Stralis-Pavese N, Tankouo Sandjong B, Bodrossy L (2006) Diagnostic microbial microarrays in soil ecology. New Phytol 171:719–736

Silva JA, Bressiani JA (2005) Sucrose synthase molecular marker associated with sugar content in elite sugarcane progeny. Genet Mol Bio 28:294–298

Tinker NA, Kilian A, Wight CP, Heller-Uszynska K, Wenzl P, Rines HR, Bjørnstad Å, Howarth CJ, Jannink J-C, Anderson JM, Rossnagel BG, Stuthman DD, Sorrells ME, Jackson EW, Tuvesson S, Kolb FL, Olsson O, Federizzi LC, Carson ML, Ohm HW, Molnar SJ, Scoles GJ, Eckstein PE, Bonman JM, Ceplitis A, Langdon T (2009) New DArT markers for oat provide enhanced map coverage and global germplasm characterization. BMC Genomics 10:39

Tomkins JP, Yu Y, Miller-Smith H, Fuisch DA, Woo SS, Wing RA (1999) A bacterial artificial chromosome library for sugarcane. Theor Appl Genet 99:419–424

Varshney RK, Graner A, Sorrells ME (2005) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10:621–630

Vettore AL, Da Silva FR, Kemper EL, Souza GM, Da Silva A, Ferro MIT, Henrique-Silva F, Giglioti ÉA, Lemos M, Coutinho LL, Nobrega MP, Carrer H, França SC, Bacci M Jr, Goldman MH, Gomes SL, Nunes LR, Camargo LE, Siqueira WJ, Van Sluys MA, Thiemann OH, Kuramae EE, Santelli RV, Marino CL, Targon M, Ferro JA, Silveira HC, Marini DC, Lemos EG, Monteiro-Vitorello CB, Tambor JH, Carraro DM, Roberto PG, Martins VG, Goldman GH, de Oliveira RC, Truffi D, Colombo CA, Rossi M, de Araujo PG,

Sculaccio SA, Angella A, Lima M, de Rosa VE Jr, Siviero F, Coscrato VE, Machado MA, Grivet L, Di Mauro SM, Nobrega FG, Menck CF, Braga MD, Telles GP, Cara FA, Pedrosa G, Meidanis J, Arruda P (2006) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res 13:2725–2735

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. Proc Natl Acad Sci USA 101:9915–9920

Wenzl P, Raman H, Wang J, Zhou M, Huttner E, Kilian A (2007) A DArT platform for quantitative bulked segregant analysis. BMC Genomics 8:196

Wittenberg AHJ, Van Der Lee T, Cayla C, Kilian A, Visser RGF, Schouten HJ (2005) Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. Mol Genet Gen 274:30–39

Xia L, Peng K, Yang S, Wenzl P, De Vicente C, Fregene M, Kilian A (2005) DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. Theor Appl Genet 110:1092–1098

Yang S, Pang W, Ash G, Harper J, Carling J, Wenzl P, Huttner E, Kilian A (2006) Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by Diversity Arrays Technology (DArT). Theor Appl Genet 113:585–595